STIC-ILL

1635

From: Schmidt, Mary
Sent: Thursday, May 16, 2002 2:41 PM
To: STIC-ILL
Subject: references, 09/646,825

please locate the following:

1. Bau et al, NAR 22:2811-2816 (1994)
2. Dean et al., NAR 14:2229-2240 (1986)
3. Malter et al, Science 246:664-666 (1989)
4. Shaw et al., Cell 46:659-667 (1986)
5. van Aarsen et al., Plant mol. biol. 28:513-524 (1995)
6. Wilson et al., Nature336:396-399 (1998).

MIC
QP620.N8

Thanks,
Melissa
11D05
11th floor mailbox

Volume 14 Number 5 1986 Nucleic Acids Research

mRNA transcripts of several plant genes are polyadenylated at multiple sites *in vivo*

Caroline Dean, Stanley Tamaki, Pamela Dunsmuir, Mitchell Favreau, Carol Katayama, Hugo Dooner and John Bedbrook

Advanced Genetic Sciences Inc., 6701 San Pablo Avenue, Oakland, CA 94608, USA

Received 23 December 1985; Revised 3 February 1986; Accepted 4 February 1986

ABSTRACT

We have analyzed the polyadenylation sites for the small subunit of ribulose bisphosphate carboxylase and chlorophyll a/b binding protein genes of Petunia (Mitchell) and the bronze gene of Zea mays. Sequence analysis of multiple cDNA clones revealed that polyadenylation of the transcripts occurred at either 2 or 3 sites for all three groups of genes. In the examples where 3 polyadenylation sites were detected, the middle site was the one predominantly used. Putative polyadenylation signals preceding the poly A tails diverged significantly from the animal consensus sequence AATAAA. In all the genes examined the first A residue in the poly A tail of the cDNA clones corresponded to an A residue in the homologous genomic sequence.

INTRODUCTION

In animals and viruses, RNA polymerase II terminates transcription far downstream of the DNA sequences coding for the 3' mRNA termini (1-3). Processing and polyadenylation of the mRNA then occurs predominantly at one site (1). A consensus sequence AATAAA is found, in the majority of animal genes sequenced, 10 to 33 nucleotides upstream from the site of polyadenylation (4). Mutational analysis has shown that the hexanucleotide AATAAA plays a major role in selecting the site at which processing and polyadenylation occurs (5,6). The consensus sequence AATAAA cannot be the only recognition signal determining the cleavage and polyadenylation site of a mature mRNA since it occurs unrecognized in the coding region and introns of several genes (7). There is now evidence suggesting that sequences immediately distal to the mature mRNA are also essential for viral and mammalian mRNA formation (8,9). The signal YGTGTTYY (Y=pyrimidine) is commonly present in mammalian genes (9), approximately 30bp downstream from the AATAAA sequence. Even

when this sequence is not present an overrepresentation of the tri-nucleotide TGT (now termed the G/T cluster) is often found downstream of the AATAAA sequence. The G/T cluster may therefore have a function in the RNA processing events.
Another sequence implicated in the positionning of the 3' cleavage site is CAYTG, usually found close to the site of polyA addition. This may base pair with a small nuclear RNA, U4 (10) and may implicate a function for U4 in the 3' cleavage process.
Although cleavage and polyadenylation occur predominantly at one site in the animal and viral messages studied, there are some exceptions (reviewed in 11). Examples where multiple polyadenylation sites have been observed include the bovine prolactin mRNA (12), the mouse ribosomal protein L30 mRNA (13), the hepatitis B virus surface antigen mRNA (14) and the Drosophila tropomyosin mRNA (15).
Processing and polyadenylation of plant mRNA's has not been studied in detail and it has been presumed that the processes in plants would follow the general rules proposed for animal genes. The putative polyadenylation signals identified in plant genes often diverge considerably from the animal consensus AATAAA (16,17). Experiments using plant mRNA's have frequently shown multiple protected fragments in 3' S1 protection experiments. These have generally been attributed to "breathing" artifacts that correspond to AT rich regions in the gene (18,19) leading to the conclusion that as in animal genes there is predominantly one polyadenylation site in plant genes.
We have studied the polyadenylation sites of three distinct groups of plant genes by sequence analysis of multiple cDNA clones. All three groups of genes, the small subunit (rbcS) genes of Petunia (Mitchell) ribulose bisphosphate carboxylase (RuBPCase), the chlorophyll a/b binding protein (Cab) genes of Petunia (Mitchell) and the bronze gene of Zea mays show multiple polyadenylation sites suggesting that this may be a general phenomenon for plant genes.

METHODS

Plant material

Th Petunia (Mitchell) strain is a doubled haploid produced by

anther culture from a hybrid between Petunia hybrida var Rose of Heaven and Petunia axillaris (20). The plants were grown under greenhouse conditions. RNA was isolated from the young leaves of plants approximately 10 weeks old.

The Zea mays line used in this study was a B;P1 version of the inbred W22 which expresses strong anthocyanin pigmentation in the husks and leaves. The plants were grown under greenhouse conditions. RNA was isolated from husk tissue of plants approximately 9 weeks old.

cDNA clone isolation and characterization

RNA was isolated as previously described (21). A cDNA library in λ gt10 was constructed from 10μg of petunia leaf poly A RNA using the method of Huynh et al (22). The primary library of $12x10^3$ phage (50% of which contained inserts) was amplified according to Huynh et al (22). The titre of the amplified library was $5x10^9$/ml. Plaques were transferred to nitrocellulose filters according to the method of Benton and Davis (23). The filters were hybridized with either rbcS or Cab cDNA clones or rbcS gene specific probes as described previously (21,24-27). The cDNA inserts were subcloned into M13 phage (28) and sequenced using the dideoxy sequencing method of Sanger et al (29).

A cDNA library in pBR322 (30) was constructed from 20μg of maize husk poly A RNA using the methods described in Maniatis et al (31). The cDNA was ligated into the EcoRI site of pBR322 after the addition of EcoRI linkers (N.E.Biolabs). The library of 32000 colonies (70% of which contained inserts) was transferred to nitrocellulose filters according to the method of Grunstein and Hogness (32).

The filters were hybridized with a genomic clone of the bronze gene (33). The bronze cDNA clones were subcloned into M13 phage (28) and sequenced using the dideoxy sequencing method of Sanger et al (29).

RESULTS

We have characterized transcripts from 6 plant genes. These fall into 3 groups; in group 1 are 3 genes from the petunia rbcS multi-gene family, in group 2 are 2 genes from the petunia Cab multi-gene family and in group 3 is the single copy bronze gene

from maize. Our characterization of the 3' regions of these genes has involved the sequencing of multiple, independent cDNA clones complementary to each of these 6 genes. In all, 37 cDNA clones have been sequenced.

Analysis of the polyadenylation sites of three petunia rbcS genes

The multi-gene family encoding the small subunit of RuBPCase in Petunia (Mitchell) consists of eight genes. The organization and expression of the individual genes has been described previously (21,24,25). We have analyzed the polyadenylation sites of two of these genes, SSU301 and SSU511 in detail and one other, SSU211 in less detail. The other rbcS genes have not been examined. The multiple cDNA clones used in the sequence analysis were isolated from a λ gt10 library constructed from petunia leaf RNA (21). Clones corresponding to the gene SSU301 were identified by their hybridization to a 61bp gene specific probe isolated from the 3' untranslated region of the SSU301 gene (21). Clones corresponding to the genes SSU511 and SSU211 were identified by hybridization to the cDNA clone pSSU51 (24,26) after high stringency washing and subsequent sequence analysis of the 3' untranslated tail region. Twelve cDNA clones corresponding to the gene SSU301, 13 corresponding to SSU511 and 2 corresponding to SSU211 were analyzed.

A comparison of the level of expression of the different rbcS genes (21) has previously indicated that the frequency of cloning in the λ gt10 library of cDNA clones for the different genes correlated with the level of mRNA for each gene as measured by Northern analysis. There was no evidence, therefore of any differential cloning of the cDNA clones corresponding to the different rbcS genes.

Figure 1 shows the nucleotide sequence of the 3' untranslated tail regions of three petunia rbcS genes and the polyadenylation sites of the analyzed cDNA clones. The cDNA clones for both of the genes SSU301 and SSU511 fall into groups with different polyadenylation sites. The numbers in brackets on Fig. 1 indicate the number of cDNA clones sequenced belonging to each polyadenylation group. The two cDNA clones analyzed corresponding to SSU211 also had different poly A sites. Since the poly A tails f all the cDNA clones analyzed were at least 50

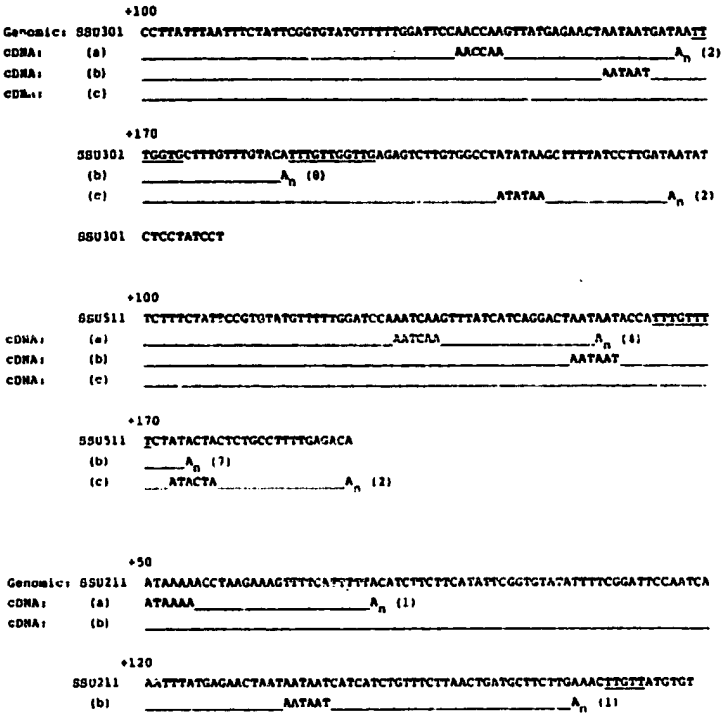

Figure 1. The nucleotide sequence of the 3' untranslated tail regions of three different petunia rbcS genes and the polyadenylation sites of multiple cDNA clones.
The position of the given nucleotide sequence relative to the translation termination codon is indicated at the beginning of each sequence. The genomic clones, SSU301 and SSU211 have been described previously (25). We have not isolated the genomic clone for SSU511, so the nucleotide sequence presented has been derived from a combination of the sequence of the cDNA clones. The cDNA clones were isolated from a λgt10 library of petunia RNA (21). cDNA clones corresponding to the gene SSU301 were identified by their hybridization to a gene specific probe (21). cDNA clones corresponding to the genes SSU511 and SSU211 were identified by hybridization to the cDNA clone pSSU51 (24,26) after high stringency washing and subsequent sequence analysis. Putative polyadenylation signals have been indicated for each group of cDNA clones. The numbers in brackets at the end of each group of cDNA clones indicate the number of cDNA clones sequenced belonging to each polyadenylation group. The sequences which may represent the G/T cluster sequences have been underlined.

nucleotides long and do not occur in a region corresponding to an AT rich area in the genomic sequence we are confident that correct priming occurred with the oligo dT during the cDNA preparation, and that the different groups of cDNA clones represent true differences in the poly A addition site on the

mRNA's. In the tw examples where many cDNA clones have been analyzed for each gene, SSU301 and SSU511, the middle example of the three polyadenylation sites is the one predominantly used. Eight out of 12 of the cDNA clones for SSU301 and 7 out of 13 clones for SSU511 had the poly A tail added in the middle position. The sites where polyadenylation occurs in the mRNA are spread over quite large distances. The three polyadenylation sites detected for SSU301 are 69 nucleotides apart, those for SSU511 are separated by 39 nucleotides; however the two detected for SSU211 are 95 nucleotides apart. A high degree of variability in the positionning of the poly A tails therefore occurs between these highly related genes.

A putative polyadenylation signal for each group of cDNA clones is illustrated in Fig.1. These were chosen as being the sequences located between 15 and 29 nucleotides upstream of the poly A tail (the consensus distance in animal genes is 10 to 33 nucleotides (4)) which most closely resemble the animal consensus sequence AATAAA (4) and which start with an A residue. One putative signal which is present in all three rbcS genes is AATAAT. This putative signal most closely resembles the animal consensus sequence and it is positionned upstream of the poly A site used predominantly in both SSU301 and SSU511. All the other putative polyadenylation signals identified diverge significantly from AATAAA.

One feature which is common to all the rbcS genes (including other genes where only one cDNA clone has been analyzed, data not shown) is that the nucleotide in the genomic sequence which corresponds to the first A in the poly A tails of the cDNA clones, is an A residue.

Also illustrated in Fig.1 are sequences which may represent the G/T cluster sequences which have been identified in animal genes (9). These are positionned immediately downstream of the poly A addition site. They are not however, present at all of the poly A addition sites, nor after all the poly A sites which are predominantly used. Their importance in plant gene mRNA formation remains questionnable. We have not found sequences corresponding to CAYTG, the sequence found in animal genes which can base pair with the small nuclear RNA, U4.

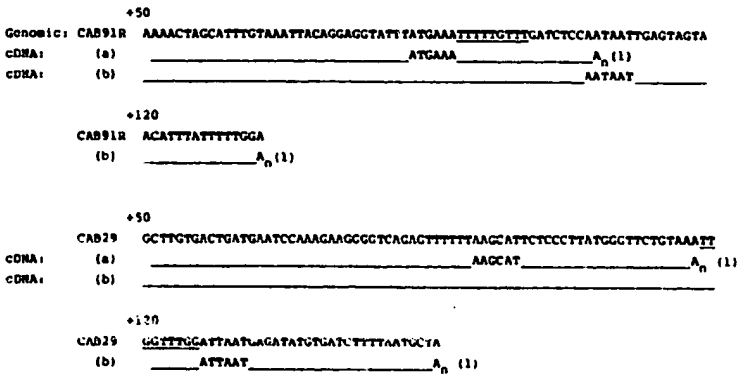

Figure 2. The nucleotide sequence of the 3' untranslated tail regions of two different petunia Cab genes and the polyadenylation sites of multiple cDNA clones.
The position of the given nucleotide sequence relative to the translation termination codon is indicated at the beginning of each sequence. The genomic clone Cab91R has been described previously (27,34). The genomic clone for Cab29 has not been isolated and the nucleotide sequence presented has been derived from a combination of the sequence of the cDNA clones. The cDNA clones were isolated by hybridization to a Cab cDNA clone, Cab 3 (27) and subsequent sequence analysis. Putative polyadenylation signals have been indicated for each group of cDNA clones. The numbers in brackets at the end of each group of cDNA clones indicate the number of cDNA clones sequenced belonging to each polyadenylation group. The sequences which may represent the G/T cluster have been underlined.

Analysis of the polyadenylation sites of two Cab genes.

The chlorophyll a/b binding proteins of Petunia (Mitchell) are also encoded by a multi-gene family consisting of at least 16 genes (27,34). We have examined the polyadenylation sites of two Cab genes by sequencing two independent cDNA clones for each gene. The cDNA clones were isolated from the λ gt10 library constructed from petunia leaf RNA by hybridization to pCab 3 probe (27), which at normal stringency hybridizes to all of the Cab genes. The cDNA clones described here were identified as corresponding to different Cab genes by subsequent sequence analysis. Figure 2 shows the nucleotide sequence of two cDNA clones for each of two different Cab genes, Cab 91R and Cab 29. Both cDNA clones for each gene have different polyadenylation sites, 28 nucleotides apart in Cab 91R and 39 nucleotides apart in Cab 29. Using the same criteria as for the rbcS genes, a putative polyadenylation signal has been pr posed for each cDNA

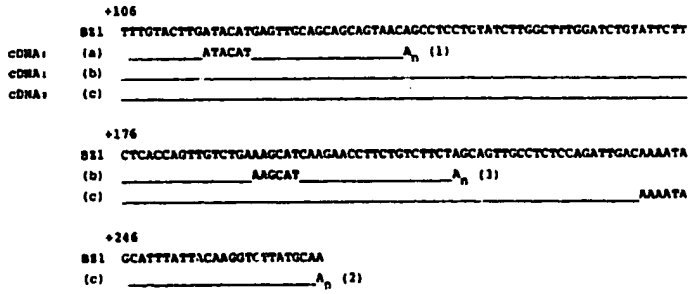

Figure 3. The nucleotide sequence of the 3' untranslated tail region of the maize bronze gene and the polyadenylation sites of multiple cDNA clones.
The position of the given nucleotide sequence relative to the translation termination codon is indicated at the beginning of the sequence. The genomic clone for the Bz allele present in the B;Pl line from which the cDNA clones were obtained has not been isolated. The bronze sequence presented has been derived from a combination of the sequence of the 6 cDNA clones and is 96% homologous to the sequence of the 3' untranslated region of the Bz-McC allele already cloned (33 and unpublished results). The cDNA clones from a pBR322 library constructed from husk tissue RNA were identified by hybridization to the genomic clone. Putative polyadenylation signals have been indicated for each group of cDNA clones. The numbers in brackets at the end of each group of cDNA clones indicate the number of cDNA clones sequenced belonging to each polyadenylation group.

clone and these are indicated in Fig. 2. Only one of the 4 putative signals in the Cab genes is AATAAT, the sequence which is present in all the rbcS genes. All other putative signals also diverge significantly from the animal sequence AATAAA (4). Where it can be determined the beginning of the poly A tail in the cDNA clones corresponds to an A residue in the genomic sequence. This is one feature common to both Cab and rbcS genes in petunia.

Analysis of the polyadenylation sites of the maize bronze gene.
The gene encoding UDP-glucose:flavonoid 3-0-glucosyltransferase (UFGT) in Zea mays is present in a single copy (33,35). Six cDNA clones for this gene were isolated from a cDNA library constructed in pBR322 from maize husk tissue RNA. The clones were identifed by hybridization to pAGS551, a 0.84kb genomic fragment corresponding to the 3' end of the transcribed region of the bronze gene (33). The 6 cDNA clones could be classified into three groups on th basis of their polyadenylati n sites. Figure

3 shows a comparison of the nucleotide sequence of the three groups of cDNA clones. The numbers in brackets on Fig.3 indicate the number of cDNA clones sequenced in each polyadenylation group. Although the sample size of cDNA clones is small it appears that as for the rbcS genes the middle polyadenylation site is the one predominantly used. This was confirmed by 3' protection experiments (data not shown). The three polyadenylation sites are separated by 130 nucleotides. Putative polyadenylation signals have been indicated on Fig.3. These were chosen using the same criteria as were described for the rbcS and Cab genes. The three polyadenylation signals chosen, ATACAT, AAGCAT and AAAATA diverge more from the animal consensus sequence, AATAAA than the polyadenylation signals identified for either the rbcS or Cab genes.

All the cDNA clones corresponding to the bronze gene have an A residue in the derived genomic sequence in the position that corresponds to the first A residue in the poly A tail. This is the one feature which has been common to all three groups of plant genes analyzed here.

DISCUSSION

In this study we have compared the polyadenylation sites for three groups of plant genes, two multi-gene families from Petunia (Mitchell), the rbcS and Cab genes and a single copy gene from Zea mays, the bronze gene. All three groups of genes show multiple sites where polyadenylation of the transcript had occurred in vivo. In the two groups of genes where many cDNA clones have been analyzed, the rbcS and bronze genes, three polyadenylation sites were detected, and in both cases the middle site is the one which is predominantly used. Nevertheless the outer polyadenylation sites are used at a frequency which is much higher than those of the alternative polyadenylation sites which have been found in some animal genes.

The data collected in this study indicate that there are not simple rules concerning the positionning of the poly A tail in plant genes. The multiple polyadenylation sites can be as close as 18 nucleotides or as distant as 130 nucleotides. They are preceded by putative polyadenylation signals which are not

Table 1. Putative polyadenylation signals identified in the rbcS, Cab and bronze cDNA clones. The putative polyadenylation consensus sequence was formulated from the 15 putative polyadenylation signals identified in the cDNA clones examined in this study.

Putative poly A signal	gene
AACCAA	rbcS
AATAAT	rbcS, Cab
ATATAA	rbcS
AATCAA	rbcS
ATACTA	rbcS
ATAAAA	rbcS
ATGAAA	Cab
AAGCAT	Cab, bronze
ATTAAT	Cab
ATACAT	bronze
AAAATA	bronze
Putative plant poly A consensus sequence:	A_{15} A_9 T_6 A_8 A_{13} T_8 T_6 A_5 C_6 T_2 A_7 G_3 T_1 C_1

conserved between genes we have analyzed, not even within a multi-gene family (summarized in Table 1). The most conserved plant polyadenylation signal is AATAAT, which is found in all the rbcS genes and one of the Cab genes. This is also the polyadenylation signal most closely resembling the animal consensus sequence, the significance of which remains unclear. A putative plant polyadenylation consensus sequence can be formulated if we compare all the putative polyadenylation signals identified in the cDNA clones examined in this study. This is outlined in Table 1. Only A or T residues are found at positions 1,2,5 and 6 in the consensus sequence.

The one general rule we can propose based on the three groups of genes analysed here is that the first A residue in the poly A tail of the cDNA clones c rresp nds to an A residue in the

homologous genomic sequence. This rule, however does not appear to hold up when other publish d plant gene sequences are included in the analysis (17). Perhaps the one general conclusion we could make is that processing and polyadenylation events for plant genes are capable of a high degree of flexibility.
It will be interesting to see if variability in the processing and polyadenylation of the plant mRNA's affects the level of mRNA stability. A preliminary comparison of the expression of the individual rbcS genes (21) with the information on their polyadenylation sites suggests there is no direct correlation. The two rbcS genes, SSU301 and SSU211, are expressed to very different levels in petunia leaf tissue, accounting for 47.3% and 1.9% respectively of the total rbcS RNA, however they both have multiple polyadenylation sites, one of which is preceded by the consensus AATAAT. Variability in the 3' processing events would not therefore appear to explain the differences in the steady state mRNA's of these two genes.

ACKNOWLEDGEMENTS

We thank Paul Brosio for the sequence analysis of the Cab cDNA clones, Edward Ralston and James English for access to the bronze genomic sequence prior to publication, Jerry Kermicle for the B;P1 maize line and Tom Lemieux for the plant material.

REFERENCES

1. Nevins, J.R. and Darnell, Jr., J.E. (1978) Cell 15, 1477-1493.
2. Citron, B., Falck-Pedersen, E., Salditt-Georgieff, M. and Darnell, Jr., J.E. (1984) Nuc. Acid Res. 12, 8723-8731.
3. Salditt-Georgieff, M. and Darnell, Jr., J.E. (1983). Proc. Natl. Acad. Sci. USA 80, 4694-4698.
4. Proudfoot, N.J. and Brownlee, G.G. (1976) Nature 263, 211-214.
5. Montell, C., Fisher, E.F., Caruthers, M.H. and Berk, A.J. (1983) 305, 600-605.
6. Wickens, M. and Stephenson, P. (1984) Science 226, 1045-1051.
7. Fitzgerald, M. and Shenk, T. (1981) Cell 24, 251-260.
8. McDevitt, M.A., Imperiale, M.J., Ali, H. and Nevins, J.R. (1984) Cell 37, 993-999.
9. McLauchlan, J., Gaffney, D., Whitton, J.L. and Clements, J.B. (1985) Nuc. Acid Res. 13, 1347-1368.
10. Berget, S.M. (1984) Nature 309, 179-182.
11. Birnstiel, M.L., Busslinger, M. and Strub, K. (1985) Cell 41, 349-359.

12. Sasavage, N.L., Smith, M., Gillam, S., Woychick, R.P. and Rottman, F.M. (1982) Proc. Natl. Acad. Sci. USA 79, 223-227.
13. Wiedemann, L.M. and Perry, R.P. (1984) Mol. Cell. Biol. 4, 2518-2528.
14. Simonsen, C.C. and Levinson, A.D. (1983) Mol. Cell Biol. 3, 2250-2258.
15. Boardman, M., Basi, G.S. and Storti, R.V. (1985) Nuc. Acid Res. 13, 1763-1776.
16. Messing, J., Geraghty, D., Heidecker, G., Hu, N.T., Kridl, J. and Rubinstein, I. (1983) in Genetic Engineering of Plants, Kosuge, T., Meredith, C.P. and Hollaender, A. Eds., pp. 211-227, Plenum Press, New York.
17. Dhaese, P., deGreve, H., Gielen, J., Seurinck, J., van Montagu, M., Schell, J. (1983) EMBO J. 2, 419-426.
18. Boutry, M. and Chua, N.H. (1985). EMBO J. 4, 2159-2165.
19. deGreve, H., Dhaese, P., Seurinck, J., Lemmers, M., van Montagu, M. and Schell J. (1982) J. Mol. Appl. Genet. 1, 499-511.
20. Mitchell, A.Z. (1979) Thesis, B.A. (Hons.) Harvard University.
21. Dean, C., van den Elzen, P., Tamaki, S., Dunsmuir, P. and Bedbrook, J. (1985) EMBO J. 4, 3055-3061.
22. Huynh, T.V., Young, R.A. and Davis, R.W. (1985) in DNA Cloning Techniques: a Practical Approach, Glover, D. Ed., pp 49-78, IRL Press, Oxford.
23. Benton, W.D. and Davis, R.W. (1977) Science 196, 180-182.
24. Dean, C., van den Elzen, P., Dunsmuir, P. and Bedbrook, J. (1985) in Molecular Form and Function of the Plant Genome, von Vloten-Doting, L., Groot, G.S.P. and Hall, T.C. Eds., pp 167-174, Plenum Press, New York.
25. Dean, C., van den Elzen, P., Tamaki, S., Dunsmuir, P. and Bedbrook, J. (1985) Proc. Natl. Acad. Sci. USA 82, 4964-4968.
26. Dunsmuir, P., Smith, S.M. and Bedbrook, J.R. (1983) Nuc. Acid Res. 11, 4177-4183.
27. Dunsmuir, P., Smith, S.M. and Bedbrook, J.R. (1983) J. Mol. Appl. Genet. 2, 285-300.
28. Messing, J., Crea, R. and Seeburg, P.H. (1981) Nuc. Acid Res. 9, 309-321.
29. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
30. Bolivar, F., Rodriquez, R.L., Greene, P.J., Betlach, M.C., Heynecker, H.L., Boyer, H.W., Crosa, J.H. and Falkow, S. (1977) Gene 2, 95-113.
31. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Laboratory, New York.
32. Grunstein, M. and Hogness, D. (1975) Proc. Natl. Acad. Sci. USA 72, 3961-3965.
33. Dooner, H.K., Weck, E., Adams, S., Ralston, E., Favreau, M. and English, J. (1985) Mol. Gen. Genet. 200, 240-246.
34. Dunsmuir, P. (1985) Nuc. Acid Res. 13, 2503-2518.
35. Federoff, N., Furtek, D., Nelson, O. (1984) Proc. Natl. Acad. Sci. USA. 81, 3825-3829.